DOCUMENT RESUME

ED 449 759 HE 033 764

AUTHOR Williams, Robert L.
TITLE Course Evaluations: A Strategy for Improving Instruction.
PUB DATE 2001-00-00
NOTE 17p.
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Course Evaluation; *Educational Improvement; Grades (Scholastic); Higher Education; Student Attitudes; *Student Evaluation of Teacher Performance; *Teaching Methods; *Undergraduate Students

ABSTRACT

A study was conducted to determine how student grades and student perceptions of instructor style affected their overall evaluations of the target course. This was part of a larger effort to study how student evaluations can be used to improve instruction. Participants were 258 students from 5 sections of an undergraduate human development course with content and course structure standardized across the sections. Following completion of a comprehensive examination and the receipt of instructor feedback about their performance on the examination, students had the necessary information to compute their final grades in the class. They were then asked to respond to a course evaluation form, using an identification number that allowed the pairing of their evaluation and grade. Correlations between grades and total course evaluation scores were statistically significant but low in magnitude. Students who received an "A" tended to rate the course higher than those who made lower grades. How students perceived the instructional style of instructors was strongly linked to their composite course evaluation. Ratings for individual aspects of the course yielded varied results. (Contains 12 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 449 759

Running head: COURSE EVALUATIONS IN INSTRUCTIONAL IMPROVEMENT

Course Evaluations: A Strategy for Improving Instruction

Robert L. Williams

The University of Tennessee

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

R. Williams

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Address:

Dr. Robert L. Williams

Department of Educational Psychology

The University of Tennessee

Knoxville, TN 37996-3400

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

HE033764

BEST COPY AVAILABLE

Course Evaluations: A Strategy for Improving Instruction

Almost all colleges and universities use some form of student evaluations in determining salary increments and promotions for faculty (Aleamoni 1987). Because they observe the targeted professors more extensively than do the professors' colleagues and supervisors, students are assumed to be in the best position for judging how well professors function in the classroom (Howard, Conway, and Maxwell 1985). Although student ratings could be valuable input in comparing how well professors function in their instructional role, perhaps their more useful function would be to help instructors improve their specific courses.

Using student evaluations to compare professors and their courses presents several problems of interpretation. Grading standards, instructor personality, age of the instructor, and student expectations about a course are among a plethora of variables that could affect student evaluations. For example, some research has indicated that students are likely to rate instructors higher when they expect to get a high grade, have younger instructors, and have full-time faculty as instructors (Frances and Gruber 1981). Given the right mix of personality characteristics and grading standards, academically weak professors could get higher student ratings than academically strong professors.

Studies that have targeted grading standards indicate that stringent standards are negatively related to course evaluations. Krautmann and Sander (1999) claim that lenient grading standards is a principal means of improving student evaluations. Wilson (1998) likewise indicates that easy graders get higher

evaluations than do tough graders. Brodie (1998) found that grade distributions across different sections of the same course were predictive of the student evaluations in those sections. Professors giving the highest grades with the least studying received the highest evaluations. Their courses were also rated as the most intellectually challenging.

Irrespective of the perceived stringency of grading standards, grades per se appear to affect student evaluations. Given access to the same videotaped lecture, students randomly assigned high grades on an exam covering the lecture rated the instructor higher than students randomly assigned lower grades (Perkins, Guerin, and Schleh 1990; Snyder and Clair 1976). The latter study found that not only did the "A" students rate the instructor's presentation more favorably, they also perceived the exam as clearer than did students assigned lower grades.

Students' perception of instructor personality or style may be another powerful contributor to students' ratings of instructor adequacy. Teachers perceived as friendly, entertaining, enthusiastic, empathic, and accommodating may receive generally favorable ratings, even when their knowledge of the subject matter is quite limited. On the other hand, professors perceived as aloof and stern may get low ratings, irrespective of their subject matter expertise. One study (Wilson 1998) reported that perceived instructor enthusiasm alone raises student evaluations.

Although student evaluations could play a legitimate role in comparing instructors and courses, a more important role would be to use them as a vehicle

for improving specific courses (Simon 1987). To accomplish the latter, a course evaluation needs to be tailored explicitly to the course being evaluated (Erwin 1994, Pulich 1984). This approach to course evaluation would not indicate necessarily the comparative ranking of an instructor, but it would provide specific information as to what aspects of a particular course may need attention.

The basic purpose of this study was to illustrate how student evaluations can be used to improve instruction. The initial objective was to determine how student grades and student perceptions of instructor style affected their overall evaluations of the target course. Instructors' concern was that perceptions of grades and instructor style would color ratings of more specific dimensions of the course. The next objective was to determine how students evaluated a variety of learning opportunities in the course. Finally, the paper examines how this explicit student feedback can be used in modifying a course.

Method

Participants

The 285 students who participated in the study came from five sections of an undergraduate human development course designed especially for students in the teacher preparation program at a large state university in the Southeast. Enrollment in the sections varied from 50 to 85 students. The total enrollment across sections was 314 students, but some students elected not to submit the evaluations on which the results of this study are based. The enrollment was predominantly sophomores (44.6%) and juniors (33.2%), although several seniors

(15.9%) and a few graduate students (6.2%) also took the course to meet credentialing requirements for the teacher preparation program. Far more females (78%) than males (22%) took the course.

Course Structure

Because the course was taught by a variety of graduate teaching assistants (GTAs) and a supervising professor, both course content and course structure were standardized across sections. Students in all sections had the same syllabus and participated in exactly the same assessment activities. All students were graded on the same criterion-referenced scale. All GTAs taught from the same set of class notes developed by the supervising instructor. The supervising professor and GTAs met weekly to monitor implementation of the course plan.

In addition to having specified reading materials and videotapes for in-class viewing, students also purchased a study guide with questions covering all content addressed in the course (including reading materials, videotapes, and instructor presentations in class). The study guide was a document of about 150 pages that highlighted all of the critical content in the course. Proportional space was left for answering the questions in this study guide, thus permitting students to take all of their notes in the study guide. Prior research (Worth 2000) regarding this course has indicated that the level of notetaking in the study guide was the one best predictor of performance on most course assessment measures.

The course was organized in units around five developmental themes: physical, cognitive, psychological, social, and moral. The course syllabus specified

what the students were to read for each unit and what was scheduled for each class session. Each class met twice weekly on Tuesday and Thursday for an hour and 15 minutes. Class sessions generally combined lecturing and student discussion. The performance measures in the course were brief essay quizzes for the five units, extensive multiple-choice exams over the units and the course as a whole, and a course project on a topic of the student's choice. Student performance was evaluated on a criterion-referenced basis.

The brief essay quiz for each unit was scheduled for the class session prior to the extensive multiple-choice exam. The essay quiz posed two questions from the reading materials section of the study guide. The two questions were selected from issues that had not been discussed in class. Students could choose one of the questions to answer, but they were not permitted to use their notes in answering the question. Students were given five minutes to respond to the question of their choice. Immediately after their papers had been taken up, the instructor presented a transparency showing the correct answers to the two questions. Student answers were graded by GTAs and returned the following class session before or after students took the unit multiple-choice examination.

Students took a 40-item multiple-choice exam at the completion of each unit and a 75-item comprehensive exam at the end of the course. The exam questions were closely linked to the questions in the study guide. Students received feedback on their exam performance as soon as they completed the exam. They were allowed to go over their scored answer sheet to determine what questions they had

missed. On the day prior to each unit exam, the instructor presented several practice items in class. The day following the exam, the instructor discussed the five most frequently missed items and explored in depth the rationale for the various choices on those items.

The course project permitted students to select a topic of their choice related to one of the five units in the course. Students were given a handout that identified potential topics and explicit guidelines for constructing their paper, including the weighting that would be given to each facet of the paper. A GTA was identified in each section of the course to work with the students in the development of their projects.

The course provided a variety of support services for the students. The course had its own web site, which allowed students to print all course documents and transparencies presented in class, keep track of their records on all course assessment activities, have access to additional instructor explanations of issues discussed in class that day, and communicate with other students in instructor-assigned study groups for the last two units in the course. The class sessions taught by the supervising professor were videotaped and made available for student viewing the same day as the class. Students in all sections, including those taught by GTAs, had access to the tapes in the instructional services center of the College of Education. Students who missed particular class sessions or had difficulty in taking notes in class were encouraged to view the tapes as needed. However, very few students took advantage of this option. The supervising instructors and all

GTAs also were available for e-mail exchanges and private conferences with students.

Course Evaluation

Following the completion of the comprehensive examination and receipt of instructor feedback regarding their performance on the exam, students had the necessary information to compute their final grade in the course. Following the opportunity to compute their final grade, students were asked to respond to a course evaluation form tailored to the structure of the course. Although they did not write their name on the evaluation form, students did supply an identification number that permitted the pairing of their evaluation with their actual grade in the course as well as with their expected grade (the grade computed by the student after the completion of the final exam).

Among the types of feedback requested on the evaluation form were the following: expected grade, comparison between the expected grade and their grade point average, extent to which the expected grade accurately represented the amount and quality of their learning in the course, and comparison between their time investment in the course and the time they typically invested in courses. In addition to this general feedback, students were asked to rate on a 0 through 3 scale (0 = no value, 1 = limited value, 2 = valuable, 3 = highly valuable) several facets of the course: overall content, reading materials, class presentations, study questions, course web site, practice exam items, project, essay quizzes, and exams. Then they were asked to evaluate the adequacy of the feedback procedures

for the essay quizzes, exams, and projects on a 0 through 3 scale (0 = totally inadequate, 1 = inadequate, 2 = adequate, 3 = highly adequate).

Because the instructional team had assumed that the exams might be a principal contributor to student evaluations of the course, students also were asked to rate on a 1 through 3 scale (1 = low, 2 = medium, 3 = high) various facets of the exams: match with study questions, emphasis on rote memorization, emphasis on thinking, clarity of exam items, and availability of assistance with items. Inasmuch as emphasis on rote memorization was viewed as a negative indicator of exam quality, it was reversed scored in the overall scoring of the student evaluations.

Finally, students were asked to select from a list of 12 descriptors (6 matched pairs) which ones best described the overall demeanor of the instructional team. Although the 12 descriptors were arranged in random order, the 6 matched pairs were the following: bland/enthusiastic, harsh/cordial, aloof/approachable, disorganized/organized, inconsistent/consistent, and unresponsive/responsive. Students could mark as many descriptors as they chose and could also add descriptors. The scoring of this item was the number of positive descriptors selected minus the number of negative descriptors selected.

Results

The results of the study are presented in two sections: (a) linkage between predictor variables and total course evaluation and (b) ratings given to different

aspects of the course. The latter analysis permits an assessment of what course variables were perceived as contributing most or least to the course.

Prediction of Course Evaluation

Correlations between grades (both actual and expected) and total course evaluation scores proved statistically significant but low in magnitude (.23 for actual grades and .25 for expected grades). An examination of total evaluation by actual grade level showed that A students rated the course significantly higher than did C and D students; B and C+ students also evaluated the course significantly higher than did the D students. In absolute terms, the A students for both actual and expected grades evaluated the course more highly than did any of the other grade levels.

A low grade in a course is particularly unacceptable when that low grade is perceived as inconsistent with grades typically received. The comparison between expected grades and GPA in the current study showed that students who expected low grades perceived those grades as being lower than their GPA, whereas students who expected high grades perceived those grades as being about on par with their GPA.

The item related to the personal demeanor of the instructional team was correlated highly with the total evaluation score. As previously noted, this item was scored as positive descriptors minus negative descriptors. This one item was correlated .60 with the total evaluation score. In general, the demeanor item yielded far more positive than negative endorsements. The mean number of positive

descriptors selected was 3.39 (out of a possible 6) and the mean number of negative descriptors was .17 (also out of a possible 6).

Ratings of Course Dimensions

With respect to the rating of different aspects of the course experience, the average ratings indicated that students (a) perceived their grade as slightly underestimating what they had learned; (b) invested about as much time in the course as they usually did in courses; (c) rated the overall content, reading materials, class presentations, study questions, and exams somewhere between valuable and highly valuable; (d) rated the course web site, practice exam items, course project, and essay quizzes as slightly less than valuable; (e) rated feedback for essay quizzes, exams, and projects between adequate and highly adequate for each type of assessment; and (f) rated most aspects of the exam as medium or above. The facet of the exam experience that received the highest rating was emphasis on thinking, and the facet that received the lowest rating was exam clarity. Overall, the study questions (included in the study guide) were rated as the most valuable part of the course and the essay quizzes as the least valuable part.

Discussion

The answer to the question of whether grades in the target course were related to the evaluation of the course is a qualified "yes." Students who made As in the course rated it higher than those who made lower grades. Nonetheless, the linkage between grades and course evaluations was not as pronounced as that suggested in past research. Perhaps this tempered relationship is partly a function

of the structure of the course, which was designed to produce a high level of success. Students who took advantage of all support options in the course seldom did poorly. In fact, 57% of the students made a B or better in the course.

The relatively high percentage of Bs and As in the course should not be construed as evidence of lenient grading standards. In fact, the instructional team who worked with this course perceived the grading standards as among the most demanding in the teacher preparation program. This perspective was apparently shared by many of the students—especially those who made C or lower—who rated their expected grade below their GPA.

How students perceived the personal style of instructors was strongly linked to their composite course evaluation. The correlation was strong enough to suggest the possibility that asking only this one question might provide almost as much information about students' overall evaluation of a course as asking numerous questions about specific aspects of the course. Perhaps if students see professors as cordial, approachable, responsive, and enthusiastic, the students will rate the course experience highly irrespective of its academic efficacy.

Student ratings of specific aspects of the course yielded some surprises. Because many students struggled with the multiple-choice exams and periodically complained about exam items, the instructional team had speculated that the exams might be the lowest rated dimension of the course. Instead, exams were rated in the top half of course dimensions (rated as slightly above valuable). Instructors in the course have noted informally that students often express a

preference for essay tests over multiple-choice tests. Yet, the essay quizzes were the lowest rated dimension of the course. The instructional team also had viewed the course web site as offering a tremendous resource to the students, but it likewise was one of the lowest rated dimensions of the course.

In some cases, an instructor will be encouraged that students perceive a dimension of a course (such as the multiple-choice exams in the current course) more favorably than had been expected. Yet high ratings of a course feature do not indicate necessarily that all is well with that dimension. Although study questions were the most highly rated feature in the course, students who rated this feature most highly did not necessarily perform better than those who rated the study questions lower. Apparently, even though most students recognized the value of the study questions, they need more guidance in how to make the best use of them.

What does an instructor do about lower ratings of a course feature? A less favorable rating of some aspect of a course (such as the essay quizzes in the current course) does not necessarily mean that this dimension should be dropped or modified. The brief essay quizzes used in this course proved strongly related to performance on the multiple-choice exams (r = .66) and total grade in the course (r = .75). The purpose of the essay quizzes was to encourage students to complete their reading and notetaking over the course materials in each unit at least one class session prior to the unit examination. Although student evaluations suggested little appreciation of this intended purpose, the essay quizzes apparently served that

purpose well. Ratings of the quizzes might be improved by providing students a more complete rationale as to their inclusion in the course.

On the other hand, the lower ratings of the course web site may suggest a need to adjust this aspect of the course. Our speculation is that the course web site needs to be more user friendly in terms of registration and navigational features. It was noted informally that many students were slow to start using the web site and only registered for the web site after repeated prodding from the instructors. The most hopeful trend in ratings of the web site was that students who saw at least some value in it made a letter grade higher than those who saw no value in it.

To maximize the information value of student evaluation, items will often need to be made more explicit. In the case of the course web site, we need to separate the ease of access from the relevance of information provided on the web site. Despite the availability of several computer labs around the university, some students may have found it inconvenient to get access to a computer. Also, instructions for registering for the web site may have proven unwieldy for some students. With respect to the lower ratings for the clarity of exam items, we need to be clearer as to what students found unclear. Did the lower ratings for this dimension suggest that students didn't understand critical terminology used in exam items, perceived the wording as convoluted, or simply found the questions hard to answer? Just as a course needs to be revised to maximize its value, course evaluations also need to be revised from one semester to the next to maximize their utility. In most cases, that revision should go in the direction of greater specificity.

References

Aleamoni, L. M. 1987. Student rating myths versus research facts. Journal of Personnel Evaluation in Education 1: 111-119.

Brodie, D. A. 1998. Do students report that easy professors are excellent teachers? The Canadian Journal of Higher Education 28, no. 1: 1-20.

Erwin, J. S. 1994. Student evaluations: Limits and prospects. Journal of Applied Research in the Community College 2, no. 1: 49-59.

Frances, S. J. and Gruber, M. B. 1981, August. Student evaluations of psychology instructors. Paper presented at the annual meeting of the American Psychological Association, Los Angeles.

Howard, G. S., Conway, C. G., and Maxwell, S. E. 1985. Construct validity of measures of college teaching effectiveness. Journal of Educational Psychology 77: 187-196.

Krautmann, A. C. and Sander, W. 1999. Grades and student evaluations of teachers. Economics of Education Review 18, no. 1: 59-63.

Perkins, D., Guerin, D., and Schleh, J. 1990. Effects of grading standards information, assigned grade, and grade discrepancies on students' evaluations. Psychological Reports 66: 635-642.

Pulich, M. A. 1984. Better use of student evaluations for teaching effectiveness. Improving College and University Teaching 32, no. 2: 91-94.

Simon, K. 1987. Using student evaluations to improve teaching and make personnel decisions ED290377.

Snyder, C. R. and Clair, M. 1976. Effects of expected and obtained grades on teacher evaluation and attribution of performance. Journal of Educational Psychology 68: 75–82.

Wilson, R. 1998. New research casts doubt on value of student evaluations of professors. Chronicle of Higher Education 44, no. 19: A12-A14.

Worth, S. L. 2000. Critical-thinking ability as a predictor of success in a large undergraduate course. Unpublished doctoral dissertation, The University of Tennessee, Knoxville.